

# Dygraph Datasets

Andrew McCrabb, Absalat Getachew, Valeria Bertacco

The datasets provided as part of DyGraph are adapted from multiple sources. The descriptions, links, and citations for each dataset are provided in this document.

---

## Attribution

### Description

A sample of 30 days of Criteo live traffic data. Each edge corresponds to one impression to an ad displayed to a user: it was clicked on, it led to conversion, and/or it led to a conversion that was attributed to Criteo.

### Original Dataset Source

<https://ailab.criteo.com/criteo-attribution-modeling-bidding-dataset/>

### Citation

Eustache Diemert, Julien Meynet, Pierre Galland, and Damien Lefortier. “Attribution Modeling Increases Efficiency of Bidding in Display Advertising.” Proc. ADKDD. 2017

---

## Bitcoin

### Description

A “who-trusts-whom” network of people who trade using Bitcoin on the platform “Bitcoin OTC”. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin OTC rate other members in a scale of -10 (total distrust) to +10 (total trust). Each edge corresponds to a rating.

### Original Dataset Source

<https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html>

### Citation

S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. Edge Weight Prediction in Weighted Signed Networks. IEEE International Conference on Data Mining (ICDM), 2016.

S. Kumar, B. Hooi, D. Makhija, M. Kumar, V.S. Subrahmanian, C. Faloutsos. REV2: Fraudulent User Prediction in Rating Platforms. 11th ACM International Conference on Web Search and Data Mining (WSDM), 2018.

---

## Email\_EU

### Description

The incoming and outgoing emails between the members of a European research institute (anonymized). The dataset only represent communication between the institution members. It doesn't contain any outgoing or incoming messages to the rest of the world. A separate edge is created for each recipient of the e-mail. The dataset available here represents the combination of the four datasets available from the original source.

### Original Dataset Source

<https://snap.stanford.edu/data/email-Eu-core.html>

### Citation

Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. "Local Higher-order Graph Clustering." In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.

J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 2007.

---

## Higgs Twitter

### Description

The Higgs dataset has been built after monitoring the spreading processes on Twitter before, during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on 4th July 2012. The messages posted in Twitter about this discovery between 1st and 7th July 2012 are considered.

### Original Dataset Source

<https://snap.stanford.edu/data/higgs-twitter.html>

### Citation

M. De Domenico, A. Lima, P. Mougél and M. Musolesi. The Anatomy of a Scientific Rumor. (Nature Open Access) Scientific Reports 3, 2980 (2013).

---

# Hospital Ward

## Description

The temporal network of contacts between patients and/or health-care workers (HCWs) in a hospital ward in Lyon, France, from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm. The study included 46 HCWs and 29 patients.

## Original Dataset Source

<http://networkrepository.com/ia-hospital-ward-proximity.php>

## Citation

Ryan A. Rossi and Nesreen K. Ahmed, “The Network Data Repository with Interactive Graph Analytics and Visualization.” AAAI. <http://networkrepository.com>. 2015.

---

# Movies

## Description

5-star rating and free-text tagging activity from “Movie Lens”, a movie recommendation service. It contains 20,000,263 ratings and 465,564 tag applications across 27,278 movies. These data were created by 138,493 users between January 09, 1995 and March 31, 2015. This dataset was generated on March 31, 2015, and updated on October 17, 2016. Users were selected at random for inclusion. All selected users had rated at least 20 movies.

## Original Dataset Source

<https://grouplens.org/datasets/movielens/20m/>

## Citation

F. Maxwell Harper and Joseph A. Konstan. “The MovieLens Datasets: History and Context.” ACM Transactions on Interactive Intelligent Systems (TiiS) 2015.

---

# Music

## Description

This dataset contains a network of listeners and music from a set of 2,000 users from Last.fm online music system.

## Original Dataset Source

<https://grouplens.org/datasets/hetrec-2011/>

## Citation

Cantador, Ivan and Brusilovsky, Peter and Kuflik, Tsvi, "2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011)." RecSys. 2011.

(Note: Not associated with a specific paper in the conference)

---

# Ubuntu

## Description

A network of interactions on the stack exchange web site "Ask Ubuntu". An edge represents an interaction between users: answering a question, commenting on a question, and commenting on an answer.

## Original Dataset Source

<https://snap.stanford.edu/data/sx-askubuntu.html>

## Citation

Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. "Motifs in Temporal Networks." In Proc WSDM. 2017.

---

# Forum

## Description

This dataset is a Facebook-like forum network of an online social network. Each edge corresponds to activity between two users on the forum.

## Original Dataset Source

[http://networkrepository.com/fb\\_forum.php](http://networkrepository.com/fb_forum.php)

### Citation

Opsahl, T., "Triadic closure in two-mode networks: Redefining the global and local clustering coefficients." Social Networks. 2011.

---

## UCI\_Messages

### Description

The Facebook-like Social Network originate from an online community for students at University of California, Irvine. The dataset includes the users that sent or received at least one message.

### Original Dataset Source

<https://snap.stanford.edu/data/CollegeMsg.html>

### Citation

Pietro Panzarasa, Tore Opsahl, and Kathleen M. Carley. "Patterns and dynamics of users' behavior and interaction: Network analysis of an online community." Journal of the American Society for Information Science and Technology. 2009.